Ⅴ．知的エージェントをめぐる刑事上のリスク及びその責任帰属

劉　　仁文

人工知能（AI）という概念は，1950年代に打ち上げられたが，それ以来，AI発展と研究は二度上下し，21世紀初頭に，三度目のブームとピークを迎えた。近年，インターネットの発展及びクラウドコンピューティングの台頭につれ，AI製品が続々と登場し，AI研究も絶えず深化している。農業文明，工業文明，ネットワーク文明という三段階を上ってきた人類文明を，四段階目としての「AI文明」へ前進させようとする勢いにある。世界各国も国の戦略全体におけるAIの重要な意味を意識し始め，関連政策を相次ぎ打ち出し，AI産業の発展を促進しようとしている。中国も同じく，2017年7月，国務院より「次世代人工知能発展計画」を公表し，AIを国の戦略というレベルに引き上げ，体系的に布石を打ち，中国におけるAI発展の「3ステップ」戦略的目標を明確にした。それを受け，中国の法学理論と法律実務業界は，AIに深く興味を示し，刑法領域における研究成果も近年爆発的に増加してきた。知的エージェントが刑事上の責任主体となれるか，その刑事責任はどのような帰属方式を採用するか，知的エージェントにどのような刑罰が適用するかなどが話題となっている。

Ⅰ．知的エージェントの刑事上のリスクは「真の問題」である

人類社会の発展を後押ししてきたほかの科学技術と同じように，AIも人間を厄介且つ単調な仕事から解放し，人類社会を後押しして躍進的な変革を遂げると同時に，そこに潜んだ不確定性から，人類の生存と発展の安全にかかわる現実的，潜在的なリスクを派生し，客観的にAIの安全性に対する極めて大きな懸念を引き起こした。

一方，AIによる刑事上のリスクに刑法は介入すべきか，知的エージェントが刑事上の責任主体になれるかといった疑問をめぐり，中国の刑法学界では，

否定の考え方もないわけではない。そして，最近，刑法学界は AI に対して深い興味を示した現象について厳しく批判し，AI をめぐる法学研究において人間の知力常識に逆らう「反知性化現象」があると指摘した学者がいる。その研究者は，以下のような指摘をしている。当面，AI をめぐる学界の法学研究は実際，「偽の問題」を「真の問題」にすり替えている。AI をめぐる学者たちのいわゆる「憂患意識」は実際，無数の架空からなる幻にすぎず，他人の「話題作り」を使って自分を驚かすに過ぎない。AI をめぐる法的リスク対策につき，今までの研究は教義のない対策論に止まり，往々にして具体的なことばかり論じており，体系的になっていない。AI が刑事上の責任を負うことができると主張する観点は，根本的に，人間が刑法を策定した道具的理性に逆らい，自由刑や死刑に対して人間は肉体的心理的に苦痛を覚える一方，ロボットは痛みを感じない。罪を犯した知的エージェントに刑罰を科するのは刑法の謙抑性原則に反する(1)。

上述した批判の観点は，ある程度急進的な AI 刑法学研究に，理性的に反省するきっかけを提供したが，AI の刑事上のリスクは「偽の命題」だとし，知的エージェントの刑事責任の帰属に関する研究は「偽の問題」を「真の問題」にすり替えるものだと断言したのは，理想的ではない。なぜなら，AI 技術に潜在したマイナス効果は人間にもたらす危害と危険を無視し，「極端から極端へ」の転向であり，間違っているからである。

まず，人類社会に知的エージェントがもたらす刑事上のリスクは，事実無根の憶測とか人騒がせな言論ではなく，客観的でリアルに存在しているのである。AI の研究開発と応用においては，プログラム故障を起因とするスマートロボットによる致死・致傷事故が複数あったのは事実である。ここで，いくつか例をあげてみる。2015年，ドイツにあるフォルクスワーゲンの工場で，技術者は作業中，ロボットにつかまれて金属製の板に押し付けられ，後亡くなった「スマートロボット殺人事件」。同年の中国で，ある CAPTCHA 回避プラットフォームが AI 技術を利用してインターネットの画像認証（CAPTCHA）のストラテジーをクラックし，国民情報を不正に獲得した「全国初の AI を利用した犯罪事案」。2016年，自動走行システム搭載のテスラ車が錯認のため起こした世界初の自動運転による死亡事故。2018年，アメリカアリゾナ州で発生した

（1）　劉艶紅「AI をめぐる法学研究の反知性化批判」東方法学（2019年，第 5 号）。

ウーバーの自動走行車による歩行者を巻き込んだ初の交通死亡事故。比較的に言えば，知的エージェントが人間を侵害する事例はまだまれだが，自律知的エージェントがプログラム故障のため暴走して人間の法益を侵害する危険性が潜んでいるのは，憶測ではないと証明できる。こうしたAIに内在した危険性または脅威にいかに適切に対応し，知的エージェントによる人間法益侵害の法的責任を科学的に配分するかは，刑法をはじめとした法制度体系において，真剣に扱うべき「真の問題」である。どの刑法措置（刑罰又は治安対策）を利用するかは，AIをめぐる刑事上のリスクに対応する刑法措置の争いであり，問題解決の具体的な方法であって，AIをめぐる刑事上のリスクが「真の問題」であるかどうかの評価に及ばない。

次に，刑法は，内在的に謙抑性があり，補充的な法益保護法である。だからと言って，決してAIの客観的な危険性を見て見ぬふりをすることではない。「刑法の謙抑性原則に沿うべきだが，科学技術の高速発展による今までと全然違った衝撃を無視してはいけない。AI技術の発展，特に将来面する可能性のある強いAIに対して，うまく対応できなければ，人類滅亡の大災難に遭う恐れがある」(2)。刑法の謙抑性の要旨は，国家刑罰権の行使を規範化させ，国家刑罰権の過度な行使又は国民の社会生活に不正な侵入を防止するところにあるが，その規制範囲は小さければ小さいほど望ましいというわけではない。ある特定の危険性が客観的に存在し，しかも刑法の介入を確実に求めているなら，刑法の規制範囲を適切に拡大するのも当然である。前述したとおり，人類社会にAIがもたらす危険性は客観的に存在しているため，刑法がAIをめぐる刑事上のリスクの予防に参与することは，法益保護の要求であり，謙抑性原則に反することではないと考えられる。

最後に，知的エージェントが人間の法益を侵害することは，今までなかった新しい問題である。これにつき，もはや疑う余地はない。批判的な立場にある研究者たちは，従来の刑法教義学のパラダイムによって，従来の刑法体系のもとで，知的エージェントをめぐる刑事責任帰属問題を解決すると望んでいるが，それは，刑法教義学と刑法立法論の間にあるインタラクティブな関係を切り離してしまったのである。刑法教義学で新しい問題を解決できない場合，刑法規範の拡張と新しい教義学の構築を積極的に模索し，刑法立法者による新し

（2）　陳叙言「AIをめぐる刑事責任主体問題に関する一考察」社会科学（2019年，第3号）。

い規範の創設に理論的なサポートを提供すべきである。知的エージェントによる人間法益侵害及びその責任帰属の場合，従来の刑法教義学では比較的に自律性の弱い知的エージェントによる人間法益侵害にかかわる刑事責任帰属を基本的に解決できるが，比較的に自律性の強い知的エージェント又は，完全自律型で独自の意識を持ってデザインとプログラムングを超えた強い AI による法益侵害行為に対して，適切に評価し処置することが難しい。それが故に，建設的な「立法論」研究が切なるものとなり，新しい責任帰属モードを導入して刑事責任配分問題を解決する必要がある。この点で言うと，AI をめぐる刑法学的研究は，批判者のいわゆる「対策論」になるが，ただこうした「対策論」は新しい教義学を踏まえたもので，従来の教義学にとって必要な補充と有益な発展だと考えられる。

Ⅱ．知的エージェントの刑事責任帰属に関する意見の相違点

知的エージェントによる刑事上のリスクは二種類に分けられる。国民の個人情報を不正に収集・獲得して詐欺犯罪を実施するなど，犯罪に乱用される道具的リスクもあれば，人間のコントロールが利かなくなり人間の法益をひどく侵害するリスクもある。後者はさらに二種類に分けられる。一つは，自律意識のない弱い AI がプログラムの動作ミスのため，人間の法益を侵害するリスクである。前述したテスラ自動走行車による死亡事故では，自動走行システムを搭載したテスラ車は交差点を通過するとき，トレーラートラックに衝突したのも，プログラムの動作ミスによるものだった。プログラムは，トレーラートラックの明るくて白い部分を空（そら）の一部だと間違って認識してしまい，その上，人間の運転手も道の状況を確認しているし，もし何かあったら，手を貸してくれると勘違いしてしまったのである(3)。もう一つは，ディープラーニングのできる強い AI で，自律的な意識のもとで，プログラムのデザインを超えて，人間法益侵害「行為」を実施するリスクである(4)。

犯罪に乱用された知的エージェントに対して責任を配分する際，犯罪行為は

（3） Max Tegmark『Life 3.0：being human in the age of artificial intelligence』（浙江教育出版社，2018年）133頁。

（4） 劉憲権「AI 時代における刑事上のリスクと刑法対策」法商研究（2018年，第 1 号）。

犯人自身の犯罪意志を示したもので，知的エージェントはただ犯人に利用された特別な道具にすぎないために，犯人が知的エージェントを乱用して犯罪を実施した客観的な様態や，犯人自身の主観的な犯罪形式に基づいて，邪悪な知的エージェントを意図的に設計・製造した行為や，注意義務の懈怠で重大な欠陥のある知的エージェントを設計・製造した行為，知的エージェントを乱用して犯罪を実施した行為，合理的監督管理義務を果たしていなかったせいでひどい結果につながった行為など，プログラムの設計者，製品の製造者及び使用者（管理者）の故意又は過失犯罪に，それぞれ刑事責任を負わせるべきである。例えば，AI 外科医は患者に新しい股関節に置換する手術をしているところ，知的エージェントに機械的誤差が発生してしまったとする。人間の医者は手術のプロセスを監督する役割を果たしておらず，途中で介入して股関節置換術を完成することができなかったため，やむを得ず，手術を早めに終了し，患者は股関節置換術を再び受けることになった。本来，当該医者は監督義務を果たせば，知的エージェントの内在的リスクが現実になっていくプロセスを切断し，ひどい結果にならずに済むはずだが，結局義務を果たさず，直ちに機械的誤差に気づいて手術の途中で介入し，完成させることはできなかった。こうして，股関節置換術の不成功と，患者は再び手術しなければならなくなったことに，客観的にそうさせた当該医者に過失犯罪で刑事責任を負わせることになる。

当面，知的エージェントをめぐる刑事責任帰属の論争における焦点と難点は，もし AI 製品が「悪くなったら」，稼動している間，プログラムから逸脱し，能動的又は受動的に人間の法益を侵害したら，刑事責任をいかにして配分するのかにある。現代責任主義では，刑事責任の発生は，正当化した根拠が必要で，誰であれ自分が実施した譴責されるべき行為にしか刑事責任を負わないと強調される。自律意識を持つ知的エージェントは，そのディープラーニング能力に導かれ，自分自身を変えて環境の変化に適応していくなかで，「自律的に」「予測のできないように」ひどく人間法益侵害の「罪」を犯す可能性がある。これを背景に考えてみると，知的エージェントを研究開発し稼動していく過程で，勤勉で責任を尽くす人間主体は，あらゆる合理的な製品責任基準に達した知的エージェントを設計し製造し，知的エージェントの悪い行為と関連するが因果関係を有していない知的エージェントによるひどい人間法益侵害行為を具体的に予測することもできない。それがために，彼らに刑事責任を負わせるやり方は，現代責任主義に直接的かつ根本的に反するのである。一方，知的エージェント自身の刑事責任主体地位を認めることで，人間によるコントロー

ルがきかなくなり人間を侵害してしまう「責任の空白」を埋めることができるだろうか。

これについて，肯定的立場にたつ研究者は，以下のように指摘した。人間によるコントロールがきかなくなり，知的エージェントは人類社会に災難的な損害をもたらす可能性があるため，刑法として，前方視的な予防とリスクの防止を重んじるべきである。AI は今伝統的な刑法責任帰属体系及びその基礎をひどく揺るがしているため，刑事責任能力から出発し，又は法人犯罪主体法理を演繹し，知的エージェントの学習能力と発展に基づき，自律的な学習能力の持つスマートロボットを「第三種類の人」「人工人」と認め，独立した刑事責任主体地位を付与し，データ削除やプログラム訂正，徹底廃棄といった新しい刑罰方法を増やし，人間によるコントロールがきかなくなった知的エージェントによる法益侵害のリスクにおける刑事責任帰属問題を妥当に解決すべきであると強調した(5)。それに対して，知的エージェントが刑事責任主体になれないと否定的に主張する研究者は，電子機械運動と人間の生理運動の間に埋めることのできないギャップが存在し，金属とプラスティック，指示電極のオン / オフというようなプログラムの組み合わせから，人間としての意識が生まれないと指摘した。人間の意識の本質及び生成メカニズムに徹底した理解がない限り，記号主義にせよコネクショニズムにせよ，人間の意識を模倣することはできないと述べた。強い AI がすでに到来したとすれば，そのために刑罰の種類を増設してもさほど意味がない。肯定派研究者によるデータ削除といった「刑罰」は，刑罰としての苦痛感という本質もなければ，権利を奪う属性もないため，性質から，強い AI への刑罰とは言えず，所有者財産へのある種の制限に過ぎないと述べた(6)。

要するに，自律的な意識のない弱い AI の刑事責任主体としての地位を否認することでさほど相違はないが，強い AI が刑事責任主体となれるかどうかに違いが生じたのである。肯定派研究者は，知的エージェント自身の道義性を立脚点とし，強い AI はすでに道具的理性の範疇を超え，ディープラーニングを経て自律的意識を形成し，人間の法益を侵害する過程でその自律的な意識を実現させる特別な主体となるとの考えである。それに対して，否定派研究者は引

（5） 劉憲権，胡荷佳「AI 時代におけるスマートロボットの刑事責任能力」法学（2018年，第 1 号）。

（6） 周銘川「強い AI の刑事責任を否定する」上海政法学院学報（2019年，第 2 号）。

き続き，人類中心主義の理念を堅持し，「人間を自然界で唯一内在的価値のあるものとみなし，当たり前のことのように人間をあらゆる価値の尺度とする。人間以外のものはすべて内在的な価値がなく，道具的な価値しかないため，人間の利益のためにサービスするのは当然だ」と強調した(7)。法律とは，人間が策定し又は認可したもので，人間の行為を規範化し，人と人の社会関係を調整する国の正式な制度である。刑法とは，法益をひどく侵害する犯罪行為に罰を与えることで，個体としての自然人又は組織と国，社会の間の関係を調整するものである。一方，知的エージェントと人間は明らかに区別されており，知的エージェントが有する「知能」は人間の「智力」ではなく，ただ，人間が詳しく予測できない状況で自立的にタスクを完成できる能力に対する一種の言い方に過ぎない。知的エージェントは，刑法規範の具体的な内容を効果的に理解することもできなければ，犯罪の実施を拒絶する反対動機も形成できないため，適格な刑事責任主体とは言えないと考えられる。

Ⅲ．知的エージェントの刑事責任主体地位を認めることは望ましくない

人間によるコントロールがきかなくなり，知的エージェントが人間の法益をひどく侵害した場合，責任帰属を判断する際，知的エージェントに刑事責任主体地位を付与する肯定説には，誤りがたくさんある。知的エージェントの刑事責任主体地位を肯定するのではなく，その道具的理性を引き続き堅持すべきだと考えられる。

１．刑事責任能力を刑事責任主体地位の十分条件とみなすのは誤りである

肯定派研究者は，知的エージェントは「独立した意志」，「識別能力と制御能力」を持ち，「自分の意志を実現させるために自主的に決定し実施する」ことができるならば，若しくは「自主性」や「自我の意識」を持つならば，刑事責任主体の条件を充たすと普遍的に考えている(8)。知的エージェントが感知した物事の本当の意味をはっきり理解するかどうかはともかく，ディープラーニングのできる知的エージェントに，刑法に求められる識別能力と制御能力がある

（７）　雷毅『人と自然：道徳の追問』（北京理工大学出版社，2015年）33頁。

（８）　夏天「AIに基づいた軍事スマート武器犯罪問題を論ずる」犯罪研究（2017年，第６号）。

と認めるとしても，知的エージェントの独立した刑事責任主体地位を肯定するわけにはいかないと考えられる。

確かに，刑事責任能力とは，行為者の識別能力と自分の行為を制御する能力であり，行為者は，自分の行為の性質及びその結果を識別し，犯罪を実施するかどうか自由に選択できる場合に限って，刑事責任を負うわけである。それに対して，刑事責任主体とは，刑法に定める犯罪行為の実施者であり，刑法による処罰を受ける刑事責任者である。刑事責任能力は，刑事責任主体地位の必要条件であるが，十分条件ではない。独立した刑事責任主体地位のある主体は，必ず刑事責任能力があるのに対して，刑事責任能力のある主体は，独立した刑事責任主体地位があるとは限らない。「『識別能力と制御能力』を持つのは人間だけではない。たとえば，飼養される動物と野生の動物も前述した能力を持っているが，犯罪の主体となれないのである。前述した能力を，生物学的意味を超えた，主に人類社会の規範と関わる『識別能力と制御能力』に特定すべきだと考えられる。それと同時に，犯罪主体から，動物など人間ではない創造物を排除すべきである」(9)。肯定説は，知的エージェントと人間の特徴とをひたすら対比し，知的エージェントに刑事責任能力がある可能性を論証してきたが，刑法理論における責任主体はつまるところ，人間しかなれないという一番基礎的で核心的な要素を見落としてしまい(10)，一部分で全体を包括し，筋の通らない考え間違いであると考えられる。

2．「自律意識と意志」と「識別と制御能力」の混同

識別と制御能力を核心とする人間の考え方や，感情，知性は，人間とほかの生物の間における最も顕著な区別のひとつであり，人間としての尊厳と奥秘である。肯定派研究者は，「自律意識と意志」イコール「識別能力と制御能力」とみなし，人工知能と人間知能の間にある客観的な差異を見落としてしまい，刑法における識別能力と制御能力の実質を曲解したのである。これに対して，人工知能の権威であるスタンフォード大学の Jerry Kaplan 教授は，「大部分の AI システムと特別なロボットが人類の脳と筋肉に似ていると考えがちだ。理解はできるが，それは危険な考え方である。…（人型ロボットの）外見からも誤解を招き，ロボットのことを，本当は我々人間に似ているように思い込ん

（9） 皮勇「AI 刑事法治の基本問題」比較法研究（2018年，第 5 号）。

（10） 於衡「AI の刑法評価アプローチ：機械規制からアルゴリズム規制」人民法治（2019年，第17号）。

で，更に我々人間社会の風習を理解のうえ従うと仮説を立てた」(11)と忠告した。社会性のある存在として，人間は機械的・物理的に社会の交流に参与するのではない。刑法において，人間の識別と制御能力にはいずれも強い社会性と規範性がある。人間は単に「有か無か」，「存在するか否か」といった事実上の特徴を感じ取るのではなく，行為の性質及びその結果に対して規範的に識別したことを前提に，ある特定の行為を実施するかしないかと意図的に選択して，社会の交流への参与や自分自身の人格の実現，社会的な期待といったニーズを充足するのである。刑法において，刑事責任を設定し追及する目的は，刑法規範に違反する行為を譴責し，その行為主体を処罰する方法で，刑法規範を守り，罪を犯してはならないように命令するところにある。規範的な識別と制御能力を持ち，刑法規範の意味を理解する人間を対象にする場合に限って，刑事責任の目的を実現できることは明らかである。

3．法人を独立した刑事責任主体とする法理に対する不当な演繹

スマートロボットがディープラーニングを通じて自律意識と意志を形成する能力を，刑事責任能力とみなすほか，知的エージェントの刑事責任主体地位を支持する肯定説は，「法人」という主体の設定理念を参考に，現行刑法における刑事責任の主体である自然人と法人に知的エージェントを追加すると考えるようである(12)。「自然人とスマートロボットの間の最も大きな区別は，自然人に命があるのに対してスマートロボットに命がないのである。しかし，命のない法人も刑事責任の主体となれたことは，『命』とは刑事責任主体の必要条件ではないことを意味する」と強調した(13)。法人の刑事責任主体地位の取得は，当面，依然として自然人をベースにしており，自然人の責任観念の基本的な範疇を超えていない。

法人が刑事責任を負う前提として，法律によって成立した法人が「法人意志」に支配され，そのメンバーが当該法人の利益を求めて刑法に定めた犯罪行為を実施し，当該犯罪行為で得た利益は当該法人に帰属する。そこで，以下の三点において，法人と知的エージェントは明らかに区別される。一つ目は，法

(11)　Jerry Kaplan『AI 時代』李盼訳（浙江人民出版社，2017年）35頁。
(12)　馬治国，田小楚「知的エージェントに刑法適用の可能性を論じる」華中科技大学学報社会科学版（2018年，第 2 号）。
(13)　劉憲権「強いスマートロボットの刑事責任主体地位否定説への返答」法学評論（2019年，第 5 号）。

人の責任主体地位は，民法や経済行政法といった刑法に前置する法律に認められたのに対して，知的エージェントは法的人格を与えられていない。域外では，スマートロボットに法的人格を与える例はあるが，AI 科学者を含めた各界から，ある種の「話題づくり」で，科学技術分野での売名行為に過ぎないと見なされ，決して知的エージェントの法的主体地位を本格的に認めたとは言えない(14)。刑法に前置する法律に，知的エージェントの法的主体地位を認められていないため，肯定説は刑法の謙抑性に反する疑いがある。二つ目は，法人の犯罪行為は法人の意志に支配されて実施しなければならないが，「法人の意志」とは，法人の意思決定部門の調整・指揮のもと，当該法人の名義で外に向けて活動する主観的な意識と意志のことである。法人の意志は依然として自然人が主導で，当該法人のメンバーらの集中的な意識であり，当該法人のメンバーが具体的に執行する。自然人が一切関与しないまま自主的に形成したものではない。三つ目は，法人犯罪による利益は，当該法人に属するのに対して，知的エージェントは権利の主体でないため，自分の「犯罪行為」による「犯罪利益」を実際に有することもないし，自分の「犯罪行為」に応じて「刑罰」を受けることもできない。したがって，同じ無生命体である法人と知的エージェントであるが，法人のほうは独立的な刑事責任主体地位があることから，強い AI にも独立的な刑事責任主体地位を与えると主張するのは，説得力が足りないように考えられる。

4．新しい刑罰方法で刑罰の機能を果たし刑罰の目的を実現することは難しい

肯定派研究者は，知的エージェントに刑事責任主体地位があるとの主張に合わせるために，人間に適用する刑法に定める刑罰方法に従って，データ削除といった新しい刑罰方法を創出し，罪を犯した知的エージェントに対する処罰に用い，知的エージェントの自己責任を実現させようとした。一方，こうした新しい刑罰方法で刑罰の機能を果たす効果は，疑われる余地がある。「AI には，人間のような感情的動機がないため，犯罪の楽しさと刑罰の苦しさを感じられないため，刑罰を受ける適格な主体となれない」(15)。「どんな刑罰措置を設計しようと，AI という特別な存在形態のため，物理的な刑罰手段では，あるべ

(14)　謝瑋「ネットで有名なロボットのソフィアは『だれ』なのか」中国経済周刊（2018年，第 5 号）。

(15)　葉良芳「AI は適格な刑事責任主体か」環球法律評論（2019年，第 4 号）。

き処罰効果は一切出ないのである。被害者の心を慰める点から見ても，現時点，マシーンに刑罰を科することで慰められるかと言えば，それはないと思われる」(16)。実際，罪を犯した知的エージェントの「犯罪記憶」を削除した後，当該知的エージェントが二度と犯罪しないことは保証できるのか。知的エージェントのプログラムを訂正して，確かに効果的だとすれば，どうして知的エージェントを設計する際，関連プログラムを訂正しなかったのか。どうして人間を侵すことを回避する，又は人間を侵したら自動的に廃棄するといったような要求を加えておかなかったのか。知的エージェントは人間によるコントロールがきかなくなり，さらに，人間を淘汰，酷使，消滅させるかもしれないといった肯定派研究者の発想によれば，人間にはひどい犯罪をした知的エージェントを徹底的に廃棄する能力と実力があるのか。肯定派研究者は，こうした疑問を重要視してほしい。

Ⅳ．治安対策で知的エージェントによる刑事上のリスクに応じる

「責任なしに犯罪はなく，責任なしに刑罰はない。犯罪を認定する時も，刑罰を科する時も，行為者の行為にある可譴責性の有無及びその程度を根拠とすべきである(17)」。知的エージェントの場合，刑法に譴責されるのに必要な相対的意志の自由もなければ，刑法上，自分の行為を識別し制御する能力，さらに刑罰を実際に感じ取り適応する能力もないため，刑事責任主体として刑事責任を帰属させるのは困難である。そして，従来の刑事責任帰属理論によって，知的エージェントに独立的な刑事責任主体地位を付与することはできない以上，知的エージェントに刑事責任を直接的に帰属させるのをやめて，社会防衛論及びそれに基づいた治安対策に視線を向けることは，知的エージェントの刑事責任帰属問題の解決にあたり，実際的な選択となると考えられる。これに対して，刑事近代学派が主張した社会防衛論に鑑み，「科学技術社会防衛論」を革新的に打ち出した研究者がいる。「AI を初めとする科学技術製品が社会に危害をもたらすリスクに対応するとき，客観的に危害が生じたり，危険が存在したりする限り，社会へ危害行為を実施した，又は，その危険性のある AI に対し

（16）　時方「AI の刑事主体地位の否定」法律科学（2018年，第 6 号）。

（17）　馮軍「刑法における責任原則――張明楷教授との討議を兼ねて」中外法学（2012年，第 1 号）。

て，治安対策の性質を持ち技術的に削除する措置を適用すべきである。そうすると，『科学技術社会防衛論』に基づいた刑事責任は，客観的な結果責任となり，こうした刑事責任を追求する形は『技術責任論』となる(18)」。社会防衛論に基づいた治安対策制度で，リスクの管理・制御や安全保障の面において，刑罰（刑事責任帰属）よりふさわしい特別な利点があることは認めるべきである。

「犯罪行為の責任の程度によって言い渡す刑罰は往々にして，刑法で犯罪を予防する目的・想定を正確かつ完全的に達成することはできない」ため(19)，刑罰の補充となる治安対策は，社会危険性に基づき，将来，社会をひどく危害する行為の実施可能性を対象とする。治安対策は，過去，社会をひどく危害した行為に対して譴責や非難など否定的に評価せず，刑事責任能力を有することを前提とせず，犯人を非難しなければならないことも必要としない(20)。これは知的エージェントの刑事上のリスクを防止するニーズと一致する。知的エージェントは自己責任の刑事責任主体である必要がなくなり，自律的に稼動する過程における内在的なリスク及びそれによる現実的な危害に基づき知的エージェント自身に責任を負わせることは困難であるため，刑法の可譴責性に関連する考え方をやめて，知的エージェントに「技術的に削除する措置」を取り，知的エージェント自身のリスクをなくすことに専念するのである。こうした意味で，物に対する治安対策措置の範疇に入れることは可能ではないか。危険な人物を対象に実施した治安対策と異なり，物に対する治安対策は，犯罪又は社会を侵害する危険を防止するために，犯罪の発生を誘導したもの，又は過去の犯罪行為で残存した危険物に対して国で講じた直接的な予防措置であり，特定の物の社会危険性をなくすことを目指す。特定の知的エージェントは，自律的に稼動する過程で，人間をひどく侵害し，また侵害し続けていく証拠がある場合，司法機関は，専門的な技術意見を聞いたうえ，当該知的エージェントに対して，技術的に危険をなくす措置を講じることができる。治安対策を利用すれば，知的エージェントの人間法益侵害における責任帰属難題を解決し，知的エージェントに内在する技術的リスクと真正面から向き合う。こうして，知的エージェントによる人間侵害の責任をどの法律主体に帰するかわからない状況に

(18)　黄雲波「AI 時代の刑事責任主体を論じる――考え違い，立場と類型」中国応用法学（2019年，第 2 号）。

(19)　林山田『刑法通論（下）』（北京大学出版社，2012年）389頁。

(20)　曹波『刑事職業禁止制度研究』（法律出版社，2018年）125頁。

陥らないで済むし，刑罰で必要となる可譴責性も避けることができる。知的エージェントの刑事上のリスクを規制する場合，知的エージェントによる技術的リスクの防止と対応に，国家公権力が介入するために，理論的なサポートと正当性根拠を提供することになる。